Visualization of the Most Effective Features Extracted by ConvNets for Detection Tasks

Maximilian Kircher, Advisor: Faraz Saeedan

TU Darmstadt, Visual Inference

In this lab, the task was, to apply the methods presented in [1] to the network of [?].

1 Class Model Visualization

1.1 VGG

First I tried to reproduce the results from [1] with a vgg network. As examples I used the category *goose*, as an example of this category is also provided by the paper. (1 a)

The objective of the optimization is, to maximize the score for this categorization network. One naïve loss function would therefore be $-score(category)$ (1 b). One problem is, that the output of the network - i.a. the score - seems to often increase, with the input. Therefore the it is possible, that this loss function leads to chaos (1 c). A solution is an additional loss, that can be added to the loss function and ensures, that the values of the image don't become to big: Each pixel is squared and the mean over all pixel values is computed. (short: sq) (1 d-h)
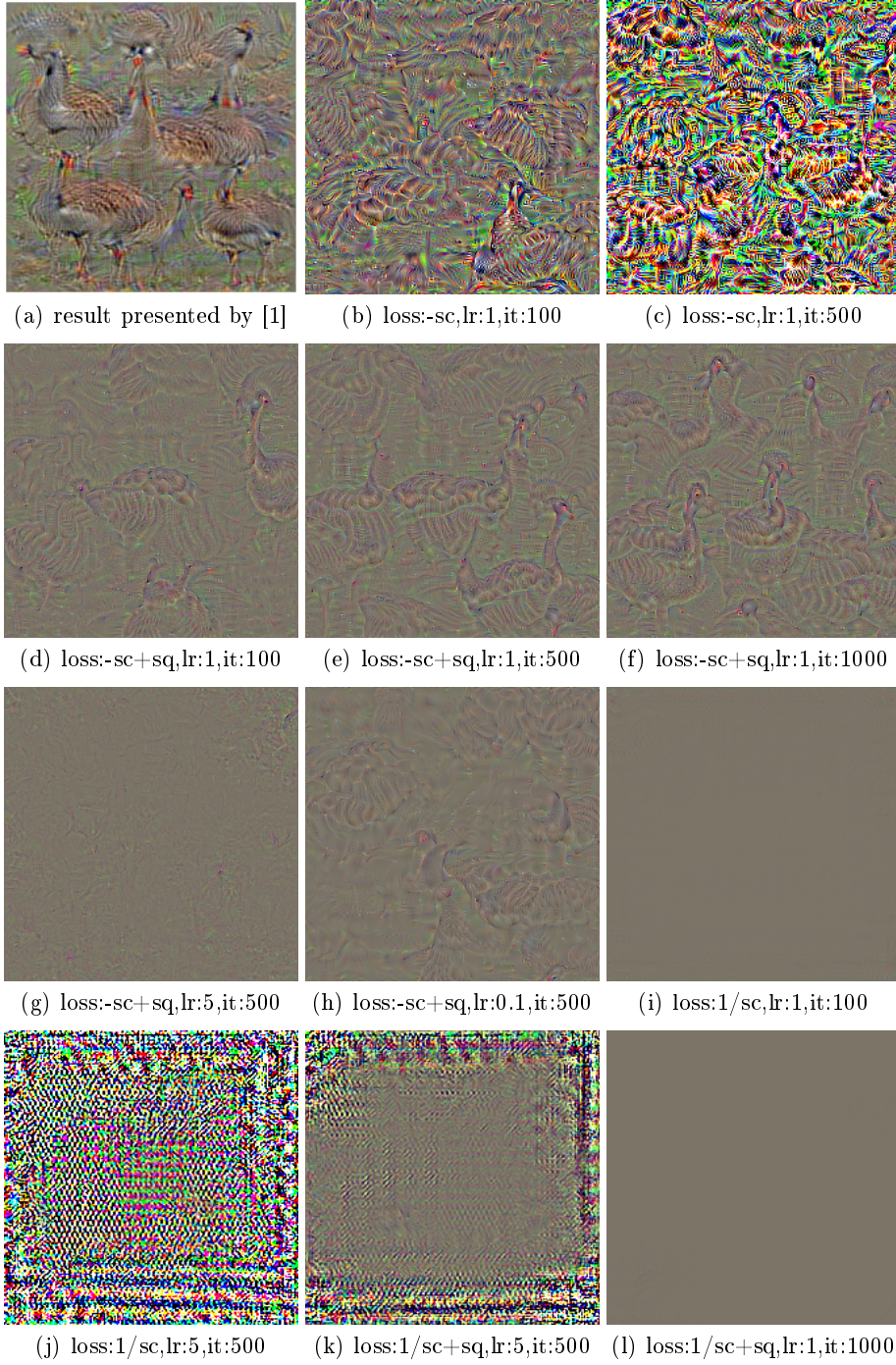
With the combined loss, the methods produces good results, that are comparable to the result of the paper.

It can be seen, that a learning rate of 1 produces good results, while the results with to high or low learning rates are not that good (1 g,h). Another loss function instead of the negative class score, that came to my mind, was $\frac{1}{score(category)}$. But experiments with this function showed, that it seems to be not useful for the task (1 i-l).

1.2 SSD

For this network I chose the category *horse*, as the implementation I used was trained for other categories, than the 1000 image net categories.

As the ssd network is an object detection network and no categorization network, the loss function, that I used for the vgg network, can not be used directly here. There are different so called prior boxes used in the network, that determine, where it looks for the objects. When I tried to use the biggest of them and maximize the category score as for the vgg network and added an additional

(a) result presented by [1] (b) loss:-sc,lr:1,it:100 (c) loss:-sc,lr:1,it:500

(d) loss:-sc+sq,lr:1,it:100 (e) loss:-sc+sq,lr:1,it:500 (f) loss:-sc+sq,lr:1,it:1000

(g) loss:-sc+sq,lr:5,it:500 (h) loss:-sc+sq,lr:0.1,it:500 (i) loss:1/sc,lr:1,it:100

(j) loss:1/sc,lr:5,it:500 (k) loss:1/sc+sq,lr:5,it:500 (l) loss:1/sc+sq,lr:1,it:1000

Figure 1. Results of the class model visualization a vgg network for the category goose. lr is the used learning rate, it the number of iterations

term, that should ensure, that the prediction for this box is localized at the whole image, it produced bad results. Also it can be seen, that it still concentrates on the upper left part.

The second loss function I tried was, to maximize the category score for all prior box prediction. In the result, can be seen, that there are very many little structures, that might be horses, but that are not that clear.

Third I used the criterion, that was also used, when training the network. It computes an localization and an confidence loss for a given target, for which I chose the wanted category and as boundaries the whole image. This loss produced the best results.

This loss I used, to visualize a smaller (1/4) object at the center of the image. There are no big differences to the bigger version...

2   Image-Specific Class Saliency Visualisation

References

1. Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.
2. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.